The Role of Reverberation and Magnitude Spectra of Direct Parts in Contralateral and Ipsilateral Ear Signals on Perceived Externalization

Song Li *, Roman Schlieper and Jürgen Peissig

Institute of Communications Technology, Gottfried Wilhelm Leibniz Universität Hannover,
DE-30167 Hannover, Germany; schlieper@ikt.uni-hannover.de (R.S.); peissig@ikt.uni-hannover.de (J.P.)
* Correspondence: song.li@ikt.uni-hannover.de; Tel.: +49-511-762-14573

check for
updates

Abstract: Several studies show that the reverberation and spectral details in direct sounds are two essential cues for perceived externalization of virtual sound sources in reverberant environments. The present study investigated the role of these two cues in contralateral and ipsilateral ear signals on perceived externalization of headphone-reproduced binaural sound images at different azimuth angles. For this purpose, seven pairs of non-individual binaural room impulse responses (BRIRs) were measured at azimuth angles of $-90°$, $-60°$, $-30°$, $0°$, $30°$, $60°$, and $90°$ in a listening room. The magnitude spectra of direct parts were smoothed, and the reverberation was removed, either in left or right ear BRIRs. Such modified BRIRs were convolved with a speech signal, and the resulting binaural sounds were presented over headphones. Subjects were asked to assess the degree of perceived externalization for the presented stimuli. The result of the subjective listening experiment revealed that the magnitude spectra of direct parts in ipsilateral ear signals and the reverberation in contralateral ear signals are important for perceived externalization of virtual lateral sound sources.

1. Introduction

Headphone-based representation of virtual acoustic environments (VAEs) is becoming more and more important due to the growing market of mobile devices, virtual and augmented reality (VR/AR) applications, high-definition television (HDTV), etc. [1]. As shown in Figure 1, the traditional way to synthesize a virtual sound source presented over headphones is by convolving a dry audio signal with a pair of binaural room impulse responses (BRIRs, the time domain representation of binaural room transfer functions (BRTFs)), which can be expressed as [2]:

$$y_i(t) = s(t) * \text{BRIR}_i(r, \theta, \varphi, t) = \int_{-\infty}^{\infty} \text{BRIR}_i(r, \theta, \varphi, t - \tau)\, s(\tau)\, d\tau, \quad i \in \{\text{left}, \text{right}\}, \tag{1}$$

where * denotes the convolution operator. $s(t)$ and $y_i(t)$ are the input audio signal and the synthesized binaural signals, respectively. r is the source-listener distance. θ and φ denote the elevation and azimuth angles of sound sources relative to the listener, respectively. The BRIR represents the direction- and distance-dependent acoustic impulse response from a sound source to the listener's eardrums in a room, consisting of:

- the direct part, which is the impulse response from the sound source to the listeners' eardrums without room information, as described by the head-related impulse response (HRIR, the time

domain representation of the head-related transfer function (HRTF)).

- the reverberant part, which contains information about reverberation (early reflections and late reverberation) stemming from the room. In the case of an anechoic chamber, the amount of reverberation is close to zero.

Perceived externalization, i.e., out of the head [3], is one of the essential indicators for the construction of immersive acoustic environments. Over the years, some important cues related to the externalization of headphone-based sound images have been discussed and investigated.
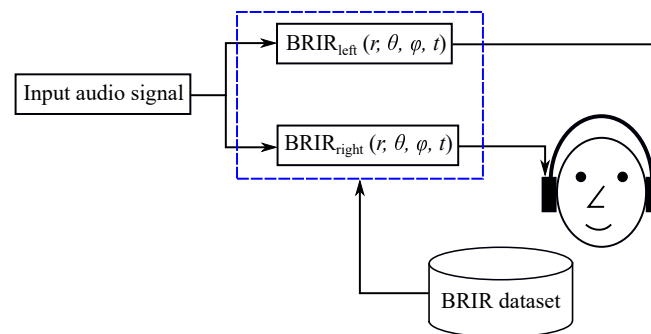

Figure 1. Illustration of the setup of BRIR measurements.

In anechoic scenarios (free-field conditions), HRTFs (direct parts in BRIRs) are responsible for perceived externalization, which depend strongly on individual anatomies, such as pinnae, head, shoulders, and torso. According to Hartmann and Wittenberg [3], Kim and Choi [4], the auditory images synthesized by using individual HRTFs could be perceived as externalized, whereas the degree of externalization was reduced if the HRTFs were non-individual. Unfortunately, recording individual HRTFs for each listener is difficult and impracticable in consumer scenarios. Due to this reason, HRTFs from some existing databases, e.g., CIPIC [5], IRCAM [6], and SADIE [7], are commonly used in dynamic binaural rendering systems. Hartmann and Wittenberg [3] showed that the interaural time differences (ITDs)/interaural phase differences (IPDs) in low-frequencies (<1 kHz), the interaural level differences (ILDs) in the full range of audible frequencies, and the monaural spectral cues from each ear were important for perceived externalization. Kulkarni and Colburn [8] have investigated the impact of reduced spectral details in HRTFs on the localization of virtual sound sources at 0°, 45°, 135°, and 180°. The magnitude spectra of HRTFs were expressed as Fourier series (FS), and the spectral details in HRTFs (magnitude spectra of HRTFs) were reduced by truncating the number of FS coefficients. The results of their study demonstrated that the performance of sound localization was unaffected by reducing the number of coefficients in the FS from 512 to 16. Furthermore, all subjects reported that the degree of perceived externalization of virtual sound images remained unchanged (informal listening test); even the number of coefficients was reduced to eight.

In non-anechoic scenarios (reverberant conditions), reverberation, especially the early reflection part, facilitates distance perception [9] and is essential for perceived externalization [10]. Several studies [11–14] concluded that the reverberation between 20 ms and 80 ms (early reflection parts) affected the degree of externalization; increasing the reverberation to durations of longer than 80 ms–100 ms did not increase externalization further. Catic et al. [15] characterized the statistical distribution of short-term binaural cues, i.e., ILDs and ITDs, in critical bands as a function of source-listener distances and investigated the impact of the sizes of these distributions (ILD and ITD fluctuations) on perceived externalization. The results of their study showed that the degree of externalization decreased substantially by compressing the ILD fluctuations of binaural speech signals containing mid- to high-frequency components (above 1 kHz). In contrast, the perceived externalization of low-pass-filtered speech sounds remained unaffected by the compression of ILD fluctuations. Therefore, the size of ILD fluctuations could be utilized as an indicator of the externalization of reverberant sound sources at mid- to high-frequencies. Catic et al. [13] pointed

out that the binaural reverberation was important for externalization, while the monaural reverberation (identical reverberation in both ears) was not sufficient to externalize a virtual sound image well, especially for frontal sound sources. This finding revealed that the reverberation-related monaural cues, i.e., direct-to-reverberant energy ratio (DRR), might by useful for perceived externalization [16], but not as relevant as the reverberation-related binaural cues, i.e., ILD fluctuations and interaural coherence (IC). Furthermore, Li et al. [14] investigated the relative influence of reverberation in contralateral versus ipsilateral ear signals on externalization of a 45° sound source. The results indicated that the reverberation at the contralateral ear had more influence on perceived externalization than that at the ipsilateral ear; the degree of externalization remained almost unchanged by removing the reverberation in the ipsilateral ear signal. However, it is still unclear how this effect changes as a source moves from lateral to frontal incidence angles.

Hassager et al. [17] studied the role of spectral details in BRIRs on perceived externalization. The magnitude spectra in the direct and reverberant part of BRIRs were smoothed by using gammatone filters with different bandwidths. The results of their study indicated that the spectral details in the direct part of BRIRs were essential for perceived externalization, while reducing the spectral details in the reverberant part of BRIRs did not affect externalization noticeably. This conclusion differs from the informal results obtained by Kulkarni and Colburn [8], where the degree of externalization was not substantially reduced by smoothing the magnitude spectra of HRTFs. It should be noted that the experiment in [8] was conducted under the free-field condition. Based on the two studies (cf. [8,17]) mentioned above, it could be concluded that the spectral details in the direct sound are important for perceived externalization of virtual sound sources when reverberation is present. However, relatively little is known about whether the spectral details in direct parts of ipsilateral and contralateral ear signals have the same influence on perceived externalization.

Other cues, e.g., dynamic cues caused by head or source movements, visual cues, etc., also play an important role in perceived externalization. Since they were not related to the present study, the details of these cues regarding externalization were not discussed in this paper. Readers can refer to some recent studies [18–22] for detailed information.

This study aimed to investigate the influence of (a) smoothing the magnitude spectra of direct parts and (b) removing the reverberation, in contralateral and ipsilateral ear signals on perceived externalization. In the studies of Catic et al. [13] and Hassager et al. [17], the reverberation and magnitude spectra of direct parts were modified for both ear signals. In order to study the influence of these two cues in contralateral and ipsilateral ear signals on perceived externalization, the reverberation and magnitude spectra of direct parts in each ear signal were manipulated separately in the present study. Li et al. [14] truncated the BRIR of each ear to different durations to study the relative influence of reverberation in contralateral versus ipsilateral ear signals on externalization of a 45° sound source. Unfortunately, there was no discussion of how this influence on externalization changed as a source moved from lateral to frontal incidence angles. In this study, the role of reverberation and magnitude spectra of direct parts in contralateral and ipsilateral ear signals was studied for virtual sound sources at various azimuth angles. The remainder of this paper is organized as follows. Section 2 introduces the modifications of direct and reverberant parts in measured BRIRs. The experimental setup for the listening experiment is described in Section 3. The subjective evaluation results are presented and discussed in Sections 4 and 5, respectively. Finally, Section 6 concludes this study.

2. Methods

2.1. BRIR Measurement

Seven pairs of BRIRs were measured with a low-noise dummy head KEMAR 45BC-12 at azimuth angles of −90°, −60°, −30°, 0°, 30°, 60°, and 90° relative to the dummy head in a listening room (area of ≈31.6 m^2, height of ≈3.2 m), which was designed under the ITU-R BS.1116 standard and has a broadband reverberation time T_{60} of about 260 ms [23]. Seven equally-calibrated Neumann KH 120A

loudspeakers located at the respective incident angles were used as sound sources for the recording of non-individual BRIRs. The distance between each loudspeaker and the dummy head was 1.9 m. A five second-long exponential sweep [24] from 20 Hz–20 kHz was used as an excitation signal, and the BRIR measurement for each azimuth angle was repeated five times. The BRIRs recorded were then truncated by a 260 ms-long time window. The measurements were performed at a sampling frequency of 44.1 kHz using a Fireface UFX+ audio interface. Figure 2 shows an illustration of the measurement setup.
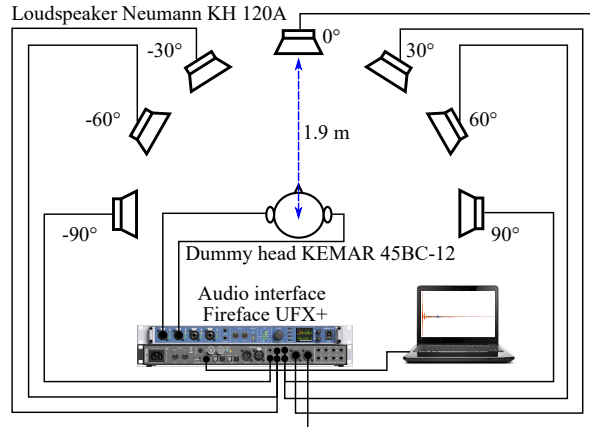

Figure 2. Illustration of the setup of BRIR measurements.

2.2. Modification of the Measured BRIR

The BRIR can be thought of as the sum of the direct ($\text{BRIR}_{\text{direct}}$) and reverberant ($\text{BRIR}_{\text{reverb}}$) part:

$$\text{BRIR}_i(r, \theta, \varphi, t) = \text{BRIR}_{\text{direct}, i}(r, \theta, \varphi, t) + \text{BRIR}_{\text{reverb}, i}(r, \theta, \varphi, t), \quad i \in \{\text{left}, \text{right}\}. \tag{2}$$

In the present study, the source-listener distance r was 1.9 m. The synthesized virtual sound sources were located in the horizontal plane ($\varphi = -90°, -60°, -30°, 0°, 30°, 60°$, and $90°$); thus the elevation angle $\theta = 0$. The direct parts in BRIRs were extracted by applying a 2.5 ms-long (110 samples at 44.1 kHz) time window with a 0.5 ms-long half raised-cosine fall time [14]. The remaining parts in BRIRs were considered as reverberant parts. The modifications of the direct and reverberant parts in BRIRs are described in the following two subsections.

2.2.1. Modification of the Direct Part in the BRIR

A previous study in [17] investigated the role of spectral details in direct parts of BRIRs on perceived externalization. For that, different levels of smoothing were achieved by using gammatone filters with various bandwidth factors B (bandwidth factors are the values relative to one equivalent rectangular bandwidth (ERB) [25]) ranging from 0.316–63.1. The smoothed spectral magnitude of the direct part, $|\text{BRTF}_{\text{direct, smooth}}(r, \theta, \varphi, f_c)|$, for each center frequency, f_c, was calculated according to Kohlrausch and Breebaart [26]:

$$|\text{BRTF}_{\text{direct, smooth}}(r, \theta, \varphi, f_c)| = \sqrt{\frac{\int_0^\infty |\text{BRTF}_{\text{direct}}(r, \theta, \varphi, f)|^2 \, |\text{H}(f, f_c)|^2 df}{\int_0^\infty |\text{H}(f, f_c)|^2 df}} \, , \tag{3}$$

where $|\text{BRTF}_{\text{direct}}(r, \theta, \varphi, f)|$ represents the spectral magnitude of the original direct part in the BRIR. $|\text{H}(f, f_c)|$ denotes the spectral magnitude of a 4$^{\text{th}}$-order gammatone filter at the center frequency of f_c with a bandwidth of $b(f_c)$, which is described in [26,27] and can be expressed as:

$$|\text{H}(f, f_c)| = \left| \left(\frac{b(f_c)}{\text{j}(f - f_c) + b(f_c)} \right)^4 \right| \, , \tag{4}$$

with

$$b(f_c) = B \times \frac{24.7\,(0.00437 \times f_c + 1)}{2\,\sqrt{2^{1/4} - 1}}, \text{ and } j = \sqrt{-1}\,. \tag{5}$$

The results showed that the externalization ratings of sound sources synthesized by such modified BRIRs (smoothed magnitude spectra of direct parts) decreased noticeably by increasing *B* above 1 ERB. In the case of the largest bandwidth factor in their study (*B* = 63.1), the synthesized binaural sound was perceived as internalized ("in the head" [28]). In the present study, 4th-order gammatone filters with *B* of 63.1 were applied to smooth "maximally" the magnitude spectra of direct parts in BRIRs. The reason for using this smoothing method is to compare the results with the observations by Hassager et al. [17]. Other methods, e.g., truncating the FS of the magnitude spectra [8], 1/N-octave smoothing [29], etc., could also be used to smooth the magnitude spectra of direct sound components.

To study the importance of spectral details in direct sounds of contralateral and ipsilateral ear signals on perceived externalization, the magnitude spectra of direct parts were smoothed (Equations (3)–(5)) in (i) BRIRs of the left ear (condition "SL"), (ii) BRIRs of the right ear (condition "SR"), and (iii) BRIRs of both ears (condition "SB"). The binaural sounds rendered by non-processed BRIRs were used as references (condition "SN"). It should be noted that in the case of −30°, −60°, and −90° sound sources, the left ear is the ipsilateral ("same-side") ear, and the right ear is the contralateral ("opposite-side") ear. In contrast, for 30°, 60°, and 90° sound sources, the left and right ear are the contralateral and ipsilateral ear, respectively. For a frontal (0°) sound source, both ears are facing the loudspeaker, and they are neither contralateral, nor ipsilateral ears.

Figure 3 shows the processed spectral magnitude of the direct part in the pair of BRIRs at 60° for the "SN" (left top), "SL" (right top), "SR" (left bottom), and "SB" (right bottom) conditions. After the smoothing process, the spectral magnitude of the direct part was almost constant across frequencies, and the specific notches and peaks disappeared. The reverberant parts in BRIRs remained unchanged and were added to the processed direct parts to generate modified BRIRs. Figure 4 shows the resulting magnitude spectra of modified BRIRs at 60° for different conditions. The magnitude spectra showed in each panel of both figures were normalized to the maximum magnitude level of the right ear BRIR.
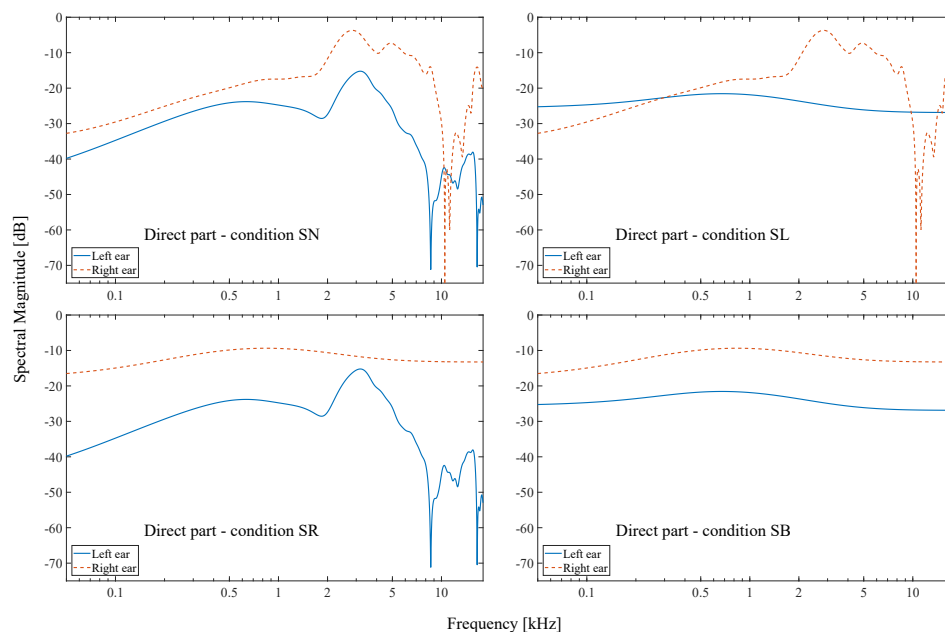

Figure 3. Examples of magnitude spectra for the direct part of the BRIR at 60° for the "SN"(left top), "SL" (right top), "SR" (left bottom), and "SB" (right bottom) conditions. The solid and dashed curves denote the magnitude spectra for the direct part in the BRIR of left and right ears, respectively.
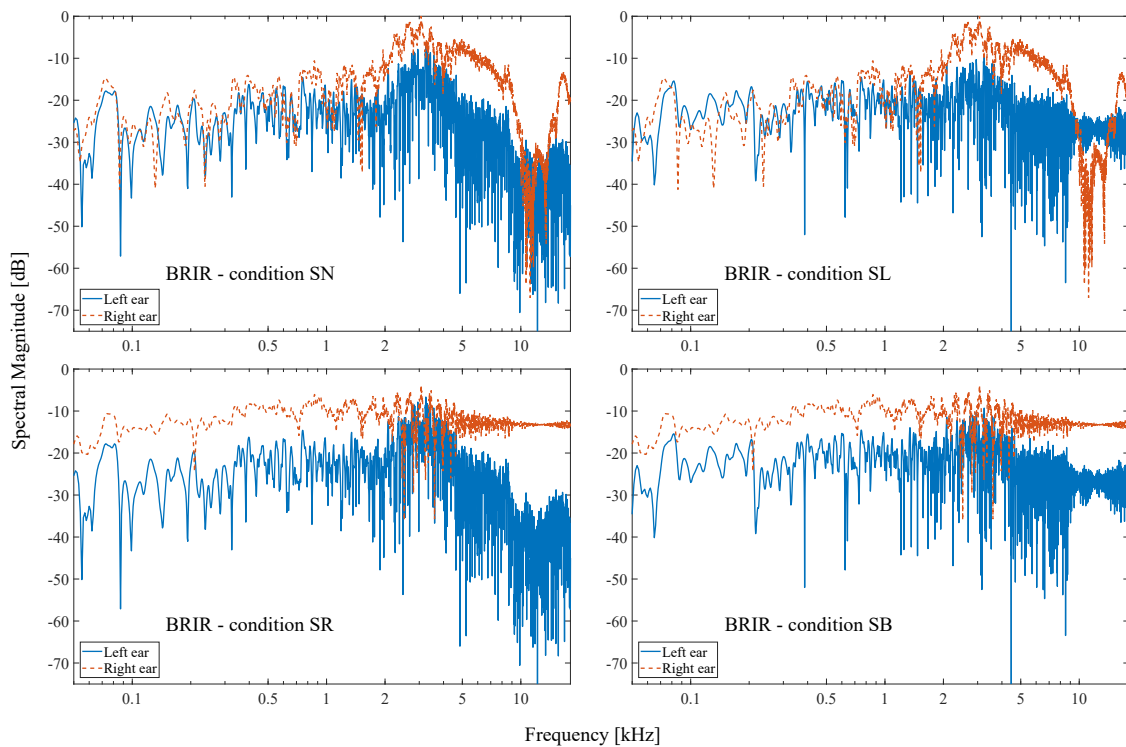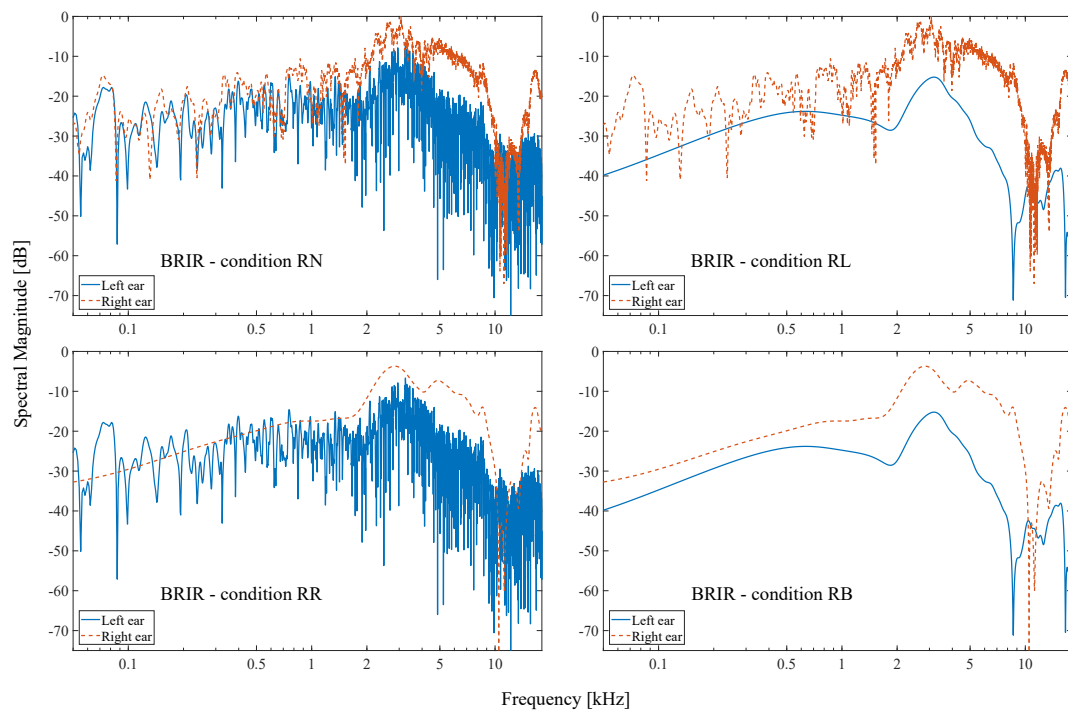
Figure 4. Examples of magnitude spectra for the BRIR at 60° for the "SN" (left top), "SL" (right top), "SR" (left bottom), and "SB" (right bottom) conditions. The solid and dashed curves denote the magnitude spectra for the BRIR of left and right ears, respectively.

2.2.2. Modification of the Reverberant Part in the BRIR

To investigate the influence of removing the reverberation in contralateral and ipsilateral ear signals on perceived externalization, the reverberant part was "maximally" removed in (i) BRIRs of the left ear (condition "RL"), (ii) BRIRs of the right ear (condition "RR"), and (iii) BRIRs of both ears (condition "RB"). The removal of reverberant parts in BRIRs was achieved by applying a 2.5 ms-long time window with a 0.5 ms-long half raised-cosine fall time (only the direct parts of BRIRs remained). All truncated BRIRs were zero-padded to a length of 260 ms. The binaural signals rendered by using non-processed BRIRs were applied as references (condition "RN"). Figure 5 shows the modified BRIRs at 60° for the "RN" (left top), "RL" (right top), "RR" (left bottom), and "RB" (right bottom) conditions in the time domain. To better visualize the BRIR modification, only the first 30 ms impulse responses are plotted, and an offset (amplitude of 1.5) was added to the left ear BRIRs for each condition. The resulting magnitude spectra for modified BRIRs at 60° are displayed in Figure 6. The magnitude spectra shown in each panel were normalized to the maximum magnitude level of the right ear BRIR. Note that the "RN" and "SN" conditions were the same, both of them denoting non-processed BRIRs.
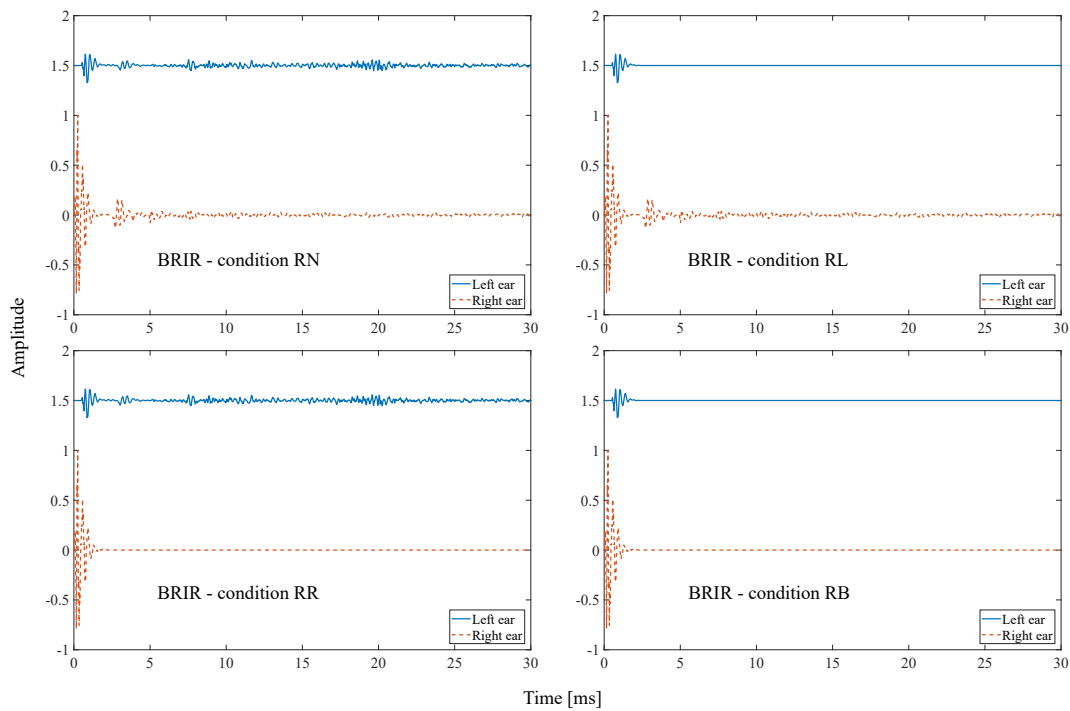
Figure 5. Examples of BRIRs with modified reverberant parts at 60° for the "RN" (left top), "RL" (right top), "RR" (left bottom), and "RB"(right bottom) conditions in the time domain. The solid and dashed curves denote the modified BRIRs of left and right ears, respectively (the left ear BRIR was offset by 1.5 for better visibility).

Figure 6. Examples of magnitude spectra for BRIRs with modified reverberant parts at 60° for the "RN" (left top), "RL" (right top), "RR" (left bottom), and "RB" (right bottom) conditions in the frequency domain. The solid and dashed curves denote the magnitude spectra for modified BRIRs of left and right ears, respectively.

3. Experimental Setup

Seven subjects (two females and five males, aged between 26 and 31) listened to the test stimuli presented over a pair of Sennheiser HD800 headphones. The effect of headphones was compensated according to Schärer and Lindau [30] by measuring the headphone transfer function (HpTF) on the dummy head (KEMAR 45BC-12). As shown in Figure 7, seven loudspeakers were placed at the measurement positions ($-90°$, $-60°$, $-30°$, $0°$, $30°$, $60°$, and $90°$ azimuth angles relative to the listener) to serve as visual cues. A subjective rating scale with four possible externalization degrees from 0–3 was used to evaluate the degree of perceived externalization, which was the same that we used in our previous study [14] (Table 1). Subjects could rate each stimulus by using a slider with a step-size of 0.1 between 0 and 3. During the listening experiment, subjects were not allowed to rotate their heads, since the perceived externalization might by reduced if listeners moved their heads without head tracking [18]. It should be noted that these two kinds of BRIR modifications were not intended to simulate naturally-occurring conditions. Rather, these artificial modifications were made to find out the relevant cues that produce the perception of externalization.

Figure 7. A photograph of the experimental setup. Seven loudspeakers were positioned at $-90°$, $-60°$, $-30°$, $0°$, $30°$, $60°$, and $90°$ relative to the subject with a distance of 1.9 m. The subject evaluated the externalization ratings of test stimuli via a graphical user interface (GUI) implemented on the Max/MSP platform [31] on a Thinkpad laptop. The Fireface UFX+ audio interface can be seen on the left side of the laptop.

Table 1. A subjective rating scale to rate the degree of perceived externalization.

Degree	Meaning of the Degree
3	The sound is externalized and at the position of the loudspeaker.
2	The sound is externalized, but not as far as the loudspeaker.
1	The sound is externalized, but very close to me.
0	The sound is in my head.

The stimulus used in the listening experiment was a truncated speech sentence (Track 50) with a length of 1.3 s taken from the European Broadcasting Union (EBU) Sound Quality Assessment Material (SQAM) [32]. For each direction, two different tests were completed by subjects, i.e., two kinds of BRIR modifications (smoothing the magnitude spectra in direct parts or removing the reverberant parts), and in each test, four audio sequences rendered by modified BRIRs, i.e., the "SL", "SR", "SB", and "SN" or "RL", "RR", "RB", and "RN" conditions, were presented in a randomized order. A repetition of each test was performed for each listener, and the externalization rating for

each experimental condition was taken as the mean of the listener's two scores. In total, 112 audio sequences (including the repetition) should be rated by each subject. Before the formal listening test, each subject was asked to listen to all stimuli once to become familiar with the perception of these stimuli. The binaurally-rendered speech signal with the non-processed BRIR was presented at a level of 64 dBA over the headphones (measured on KEMAR). During the listening test, listeners were able to repeat every sequence. They could listen to the original stimulus played back through loudspeakers and were informed that such stimulus should act as a "fully-externalized" signal (externalization rating = 3). Each listener took about 55 min to complete the whole experiment. This experiment aimed to evaluate the degree of perceived externalization of presented stimuli. Therefore, other perceptual attributes, e.g., coloration [12], plausibility [33], the accuracy of source localization [34], etc., were not evaluated.

4. Results

4.1. Smoothing the Magnitude Spectra of Direct Parts in BRIRs

Figure 8 shows the mean externalization ratings of the test stimuli for the "SN" (diamonds), "SL" (triangles), "SR" (circles), and "SB" (squares) conditions in different azimuth angles. In general, the externalization ratings increased as the sound source moved from frontal to lateral incidence angles for all conditions. In the case of non-processed BRIRs ("SN" condition), the overall mean externalization ratings were high across different azimuth angles (between 2.2 and 2.8), corresponding to the sound sources being well externalized and very close to the loudspeakers' positions and were clearly higher than those for the other three ("SB", "SL", and "SR") conditions.
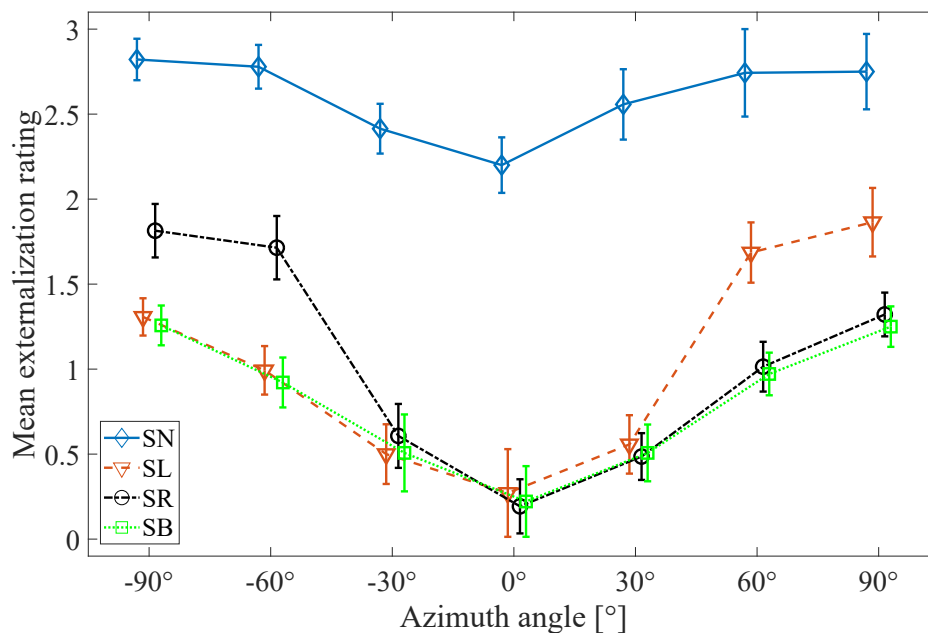
Figure 8. Mean externalization ratings from zero (the sound is in my head) to three (the sound is externalized and at the position of the loudspeaker) across listeners for the "SN" (diamonds), "SL" (triangles), "SR" (circles), and "SB" (squares) conditions in different azimuth angles. The error bars indicate the standard deviations of the mean values.
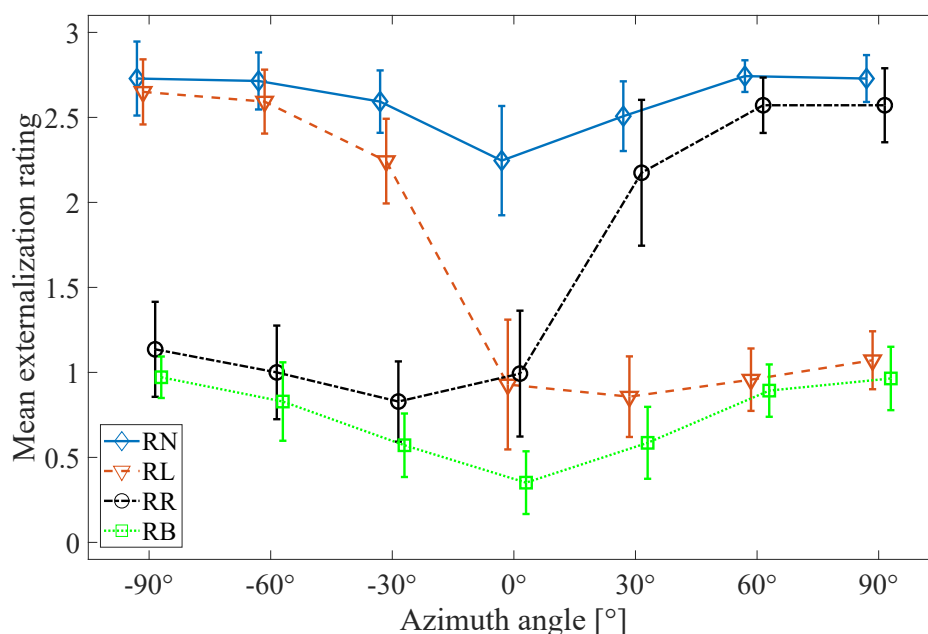
For the "SB" condition, the mean externalization ratings were below 1.0 for azimuth angles of 0° and ±30°, corresponding to sound sources being fully internalized ("in the head"). The sound images were perceived as close to subjects' heads or slightly externalized for ±60° and ±90° sound sources. For azimuth angles of −60° and −90°, the externalization ratings were higher for the "SR" condition than the "SB" and "SL" conditions, while the externalization ratings for the "SL" condition

were higher than those for the "SB" and "SR" conditions in the case of 60° and 90° sound sources. In addition, the externalization ratings for the "SL" and "SB" conditions were almost the same in the case of negative azimuth angles, while the ratings for the "SR" and "SB" conditions were almost identical for positive azimuth angles. No substantial difference in externalization ratings was observed among the three ("SL", "SR", and "SB") conditions for azimuth angles of 0° and ±30°.

For each azimuth angle, an analysis of variance (ANOVA) with post-hoc-test (5% significance level with Bonferroni adjustment) was performed to show the significant differences of externalization ratings among the "SN", "SL", "SR", and "SB" conditions. Overall, consistent with the visual inspection of Figure 8, the results for the "SN" condition differed significantly from the results for the other three conditions across various azimuth angles ($p \ll 0.05$). In the case of azimuth angles of 0° and ±30°, no significant difference was observed among the "SL", "SR", and "SB" conditions ($p > 0.2$). For lateral sound sources located at −60° and −90°, the results for the "SR" condition differed significantly from the results for the "SL" and "SB" conditions ($p \ll 0.05$), while there was no significant difference between the externalization ratings for the "SL" and "SB" conditions ($p > 0.3$). In the case of 60° and 90° sound sources, there was no significant difference between the externalization ratings for the "SR" and "SB" conditions ($p > 0.3$), but each differed significantly from the results for the "SL" condition ($p \ll 0.05$).

4.2. Removing the Reverberant Parts in BRIRs

Figure 9 shows the mean externalization ratings of the test stimuli for the "RN" (diamonds), "RL" (triangles), "RR" (circles), and "RB" (squares) conditions in different azimuth angles. For non-processed BRIRs ("RN" condition), the mean externalization ratings were overall high across different azimuth angles and highly consistent with the ratings for the "SN" condition. The degree of perceived externalization reduced substantially by removing the reverberant parts in BRIRs of both ears for all azimuth angles ("RB" condition).

Figure 9. Mean externalization ratings from zero (the sound is in my head) to three (the sound is externalized and at the position of the loudspeaker) across listeners for the "RN" (diamonds), "RL" (triangles), "RR" (circles), and "RB" (squares) conditions in different azimuth angles. The error bars indicate the standard deviations of the mean values.

In the case of the "RB" condition, the mean externalization ratings increased as the sound source moved from frontal to lateral incidence angles, but they were overall lower than 1.0, corresponding

to sound sources being internalized or very close to subjects' heads. In addition, the ratings were lower than those for the other three conditions for all azimuth angles. For a frontal sound source, the externalization ratings were almost the same for the "RL" and "RR" conditions. In the case of negative azimuth angles, the externalization ratings were substantially higher for the "RL" condition than the "RB" and "RR" conditions and slightly lower than the "RN" condition. The ratings for the "RR" condition were a little higher than those for the "RB" condition. In contrast, for positive azimuth angles, the ratings were higher for the "RR" condition than the "RB" and "RL" conditions and a little lower than the "RN" condition. Furthermore, the ratings were slightly higher for the "RL" condition than the "RB" condition.

An ANOVA with post-hoc-test (5% significance level with Bonferroni adjustment) was performed to show the significant differences of externalization ratings among the "RN", "RL", "RR", and "RB" conditions for each azimuth angle. In the case of the 0° sound source, no significant difference was observed between the externalization ratings for the "RL" and "RR" conditions ($p = 0.8$), but each of them differed significantly from the results for the "RB" and "RN" conditions. A significant difference was found between the results for the "RN" and each of the other three ("RL", "RR", and "RB") conditions ($p \ll 0.05$). For negative azimuth angles, there was generally no significant difference between the externalization ratings for the "RN" and "RL" conditions ($p > 0.2$), apart from one exception at the azimuth angle of $-30°$, where externalization was significantly higher for the "RN" condition than the "RL" condition ($p = 0.03$). However, the results for both conditions differed significantly from the results for the "RR" and "RB" conditions. Furthermore, there was no significant difference between the "RR" and "RB" conditions ($p > 0.2$). For positive azimuth angles, no significant difference was found between the externalization ratings for the "RN" and "RR" conditions ($p > 0.2$), and the results for both conditions differed significantly from the results for the "RL" and "RB" conditions ($p \ll 0.05$). Furthermore, there was no significant difference between the "RL" and "RB" conditions ($p > 0.5$).

5. Discussion

Previous studies in [13,17] revealed that the spectral details in direct parts of BRIRs and the degree of reverberation (length of BRIRs) were two important cues for perceived externalization of virtual sound sources in reverberant environments. Either smoothing the magnitude spectra of direct parts or truncating the reverberant parts of BRIRs, the perceived externalization decreased substantially. In the present study, the importance of these two cues in ipsilateral and contralateral ear signals for perceived externalization was further investigated.

As mentioned in Section 1, non-individual BRIRs are commonly used in binaural rendering systems because the measurement of individual BRIRs is not feasible for each listener in every room in normal consumer scenarios [18,20]. To generalize display scenarios, non-individual BRIRs were used in our study to synthesize binaural signals. It can be seen that the simulated virtual sound sources rendered by original BRIRs ("SN" and "RN" conditions) were all perceived as externalized (mean externalization ratings > 2.0). The mean externalization rating was lower for a frontal sound source than a lateral sound source and generally increased as the source moved from frontal to lateral incidence angles, which was consistent with the results of the studies in [4,35,36].

5.1. The Role of Spectral Details in Direct Parts on Perceived Externalization

The results from the study in [17] revealed that the magnitude spectra in direct parts of BRIRs contributed to sound externalization in reverberant environments, while the spectral details of the reverberant parts had little impact on externalization. It can be assumed that the spectral details in the reverberation of either the left or the right ear BRIRs could not noticeably affect perceived externalization. Therefore, our study focused on the role of the magnitude spectra of direct parts in contralateral and ipsilateral ear signals on perceived externalization.

The experimental results of the "SN" and "SB" conditions confirmed the conclusions reported in [17], i.e., the magnitude spectra in the direct parts of BRIRs were important to perceived externalization, the degree of externalization decreased significantly by smoothing the magnitude spectra in BRIRs of both ears. It can be observed that the degree of externalization decreased significantly by smoothing the magnitudes spectra of direct parts in either left or right ear signals. This means that an externalized sound source required complete spectral details in direct parts of both ear signals; maintaining the correct magnitude spectra in direct sounds only of the left or the right ear was not sufficient to externalize a sound image well.

As a source moved from frontal to lateral incidence angles, the relative influence of spectral details in direct parts of contralateral versus ipsilateral ear signals on perceived externalization could be observed. Again, the contralateral ear signals denote the left and right ear signals for the positive and negative azimuth angles, respectively. In contrast, the ipsilateral ear signals are the left and right ear signals for the negative and positive azimuth angles, respectively. For azimuth angles of 0° and ±30°, the spectral details in direct parts of left and right ear signals had almost the same influence on externalization. The contribution of the spectral details in direct parts of ipsilateral ear signals on perceived externalization increased as the source moved laterally. In the case of large azimuth angles of ±60° and ±90°, the externalization ratings were significantly higher for smoothed magnitude spectra in direct parts of the contralateral ear signals than those of the ipsilateral ear signals.

During this experiment, all listeners reported noticeable localization biases of sound images for the "SL" and "SB" conditions at −30° and −60° and for the "SR" and "SB" conditions at 30° and 60°, i.e., the perceived sound sources were located in ipsilateral sides of heads, very close or immediately adjacent to their ipsilateral ears. It appeared that a strong smoothing of the magnitude spectra in direct parts of ipsilateral ear signals led to the problem of source localization, which was consistent with the results from the studies in [37,38]. Since our study focused on perceived externalization, the listeners were asked only to evaluate the degree of externalization for the stimuli presented. They found that the sound source was diffuse and very close to their heads under the conditions mentioned above. None of them reported that the sound was well externalized and perceived far away from their heads. Morimoto [37], Macpherson and Sabin [38] illustrated that the contribution of the ipsilateral ear on the sound source localization increased as the source moved laterally. In contrast, the contribution of the contralateral ear decreased by increasing the source angle. For the azimuth angles larger than 60°, the contralateral ear did not have contributions to determine the source localization. This result revealed that the spectral information in the ipsilateral ear signal was more important to the sound localization than that in the contralateral ear signal. Macpherson and Sabin [38] further quantified the binaural weighting of spectral cues of each ear as a function of the azimuth angle of the stimulus. Baumgartner et al. [39] chose a sigmoid function to determine the binaural weighting of spectral information in each ear signal for the source localization as a continuous increasing function of the lateral angle.

Through the experiment conducted in this study, a similar contribution of spectral details in direct parts of the left and right ear signal to perceived externalization and sound localization was observed, i.e., the contribution of the spectral details in direct parts of ipsilateral ear signals to perceived externalization was higher than that of contralateral ear signals. However, the spectral details in direct parts of contralateral ear signals always had contributions to perceived externalization, even for azimuth angles larger than 60°; maintaining the correct magnitude spectra in direct sounds only of one ear was not sufficient to externalize a sound image well.

5.2. The Role of Reverberation on Perceived Externalization

The experimental results shown in Figure 9 confirmed the findings of our previous study in [14], i.e., the reverberation at the contralateral ear had more influence on perceived externalization than that at the ipsilateral ear. During this experiment, most listeners reported that the sound images were perceived as unnatural for the "RL" and "RR" conditions since the energy of the lateralized

reverberation was strongly changed. As mentioned before, listeners should ignore the naturalness of the stimuli, and they were only asked to evaluate perceived externalization of stimuli.

For a frontal sound source, the degree of externalization decreased significantly by removing the reverberation in either left or right ear signals. The contribution of reverberation at the contralateral ear to perceived externalization increased as the sound source moved from frontal to lateral incidence angles. For azimuth angles larger than 30°, the reverberation at the ipsilateral ear had only a slight contribution to the externalization of sound images, and the reverberation at the contralateral ear dominated the judgment of perceived externalization. This finding may be utilized in the design of binaural rendering systems. In the case of lateral sound sources (azimuth angles larger than 30°), the amount of reverberation at the ipsilateral ear can be reduced appropriately to reduce the computational complexity, while taking into account the perceptual attributes such as listener envelopment, naturalness, etc.

6. Conclusions

The present study investigated the role of reverberation and magnitude spectra of direct parts in contralateral and ipsilateral ear signals on perceived externalization of virtual sound sources at different azimuth angles in a listening room. Two modifications were performed on the measured BRIRs of each ear, (a) smoothing the magnitude spectra of the direct parts and (b) removing the reverberant part. Such modified BRIRs were convolved with a 1.3 s-long speech signal, and the resulting binaural sounds were used for the listening experiment. The result of the listening experiment revealed the following:

(1) Spectral details in direct parts of the ipsilateral ear signals were more important for perceived externalization than those of the contralateral ear signals.

(2) The contribution of the spectral details in direct sounds of the ipsilateral ear signals to perceived externalization increased as the source moved laterally. However, maintaining the correct magnitude spectra in direct sounds only at the ipsilateral ear was not sufficient to externalize a sound image well.

(3) The reverberation at the contralateral ear had more influence on perceived externalization than that at the ipsilateral ear.

(4) The contribution of the reverberation at the contralateral ear to perceived externalization increased as the source moved laterally. For azimuth angles larger than 30°, the reverberation at the ipsilateral ear did not have noticeable contributions to externalize sound images.

Further work in our framework is to relate perceived externalization to physical binaural signals and to set up an externalization model to objectively predict the degree of externalization using the ILD-based [17], dynamic binaural cues-based (ILD and IC fluctuation) [14], and the monaural cues-based [40] concepts.

Abbreviations

The following abbreviations are used in this manuscript:

VAE   Virtual acoustic environment
VR   Virtual reality
AR   Augmented reality
HDTV   High-definition television
BRIR   Binaural room impulse response
BRTF   Binaural room transfer function
HRIR   Head-related impulse response
HRTF   Head-related transfer function
FS   Fourier series
ITD   Interaural time difference
IPD   Interaural phase differences
ILD   Interaural level difference
DRR   Direct-to-reverberant energy ratio
IC   Interaural coherence
ERB   Equivalent rectangular bandwidth
HpTF   Headphone transfer function
EBU   European Broadcasting Union
SQAM   Sound quality assessment material
ANOVA   Analysis of variance

References

1. Sunder, K.; He, J.; Tan, E.L.; Gan, W.S. Natural sound rendering for headphones: integration of signal processing techniques. *IEEE Signal Process. Mag.* **2015**, *32*, 100–113. [CrossRef]
2. Vorländer, M. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
3. Hartmann, W.M.; Wittenberg, A. On the externalization of sound images. *J. Acoust. Soc. Am.* **1996**, *99*, 3678–3688. [CrossRef] [PubMed]
4. Kim, S.M.; Choi, W. On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. *J. Acoust. Soc. Am.* **2005**, *117*, 3657–3665. [CrossRef] [PubMed]
5. Algazi, V.R.; Duda, R.O.; Thompson, D.M.; Avendano, C. The cipic hrtf database. In Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 21–24 October 2001; pp. 99–102.
6. Listen HRTF Database. Available online: http://recherche.ircam.fr/equipes/salles/listen/ (accessed on 15 December 2018).
7. Kearney, G.; Doyle, T. An HRTF database for virtual loudspeaker rendering. In *Audio Engineering Society Convention 139*; Audio Engineering Society: New York, NY, USA, 2015.
8. Kulkarni, A.; Colburn, H.S. Role of spectral detail in sound-source localization. *Nature* **1998**, *396*, 747. [CrossRef] [PubMed]
9. Shinn-Cunningham, B. Learning Reverberation: Considerations for Spatial Auditory Displays. In Proceedings of the 6th International Conference on Auditory Display (ICAD), Atlanta, GA, USA, 2–5 April 2000.
10. Begault, D.R.; Wenzel, E.M.; Anderson, M.R. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.* **2001**, *49*, 904–916. [PubMed]
11. Völk, F. Externalization in data-based binaural synthesis: effects of impulse response length. In Proceedings of the International Conference on Acoustics (NAG/Tagungsband Fortschritte der Akustik-DAGA 2009), Rotterdam, The Netherlands, 23–26 March 2009; pp. 1075–1078.
12. Crawford-Emery, R.; Lee, H. The Subjective Effect of BRIR Length on Perceived Headphone Sound Externalization and Tonal Coloration. In *Audio Engineering Society Convention 136*; Audio Engineering Society: Berlin, Germany, 2014.

13. Catic, J.; Santurette, S.; Dau, T. The role of reverberation-related binaural cues in the externalization of speech. *J. Acoust. Soc. Am.* **2015**, *138*, 1154–1167. [CrossRef] [PubMed]

14. Li, S.; Schlieper, R.; Peissig, J. The effect of variation of reverberation parameters in contralateral versus ipsilateral ear signals on perceived externalization of a lateral sound source in a listening room. *J. Acoust. Soc. Am.* **2018**, *144*, 966–980. [CrossRef] [PubMed]

15. Catic, J.; Santurette, S.; Buchholz, J.M.; Gran, F.; Dau, T. The effect of interaural-level-difference fluctuations on the externalization of sound. *J. Acoust. Soc. Am.* **2013**, *134*, 1232–1241. [CrossRef] [PubMed]

16. Zahorik, P.; Brungart, D.S.; Bronkhorst, A.W. Auditory distance perception in humans: A summary of past and present research. *ACTA Acust. United Acust.* **2005**, *91*, 409–420.

17. Hassager, H.G.; Gran, F.; Dau, T. The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. *J. Acoust. Soc. Am.* **2016**, *139*, 2992–3000. [CrossRef] [PubMed]

18. Hendrickx, E.; Stitt, P.; Messonnier, J.C.; Lyzwa, J.M.; Katz, B.F.; De Boishéraud, C. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *J. Acoust. Soc. Am.* **2017**, *141*, 2011–2023. [CrossRef] [PubMed]

19. Hendrickx, E.; Stitt, P.; Messonnier, J.C.; Lyzwa, J.M.; Katz, B.F.; De Boishéraud, C. Improvement of Externalization by Listener and Source Movement Using a "Binauralized" Microphone Array. *J. Audio Eng. Soc.* **2017**, *65*, 589–599. [CrossRef]

20. Li, S.; Schlieper, R.; Peissig, J. The Impact of Trajectories of Head and Source Movements on Perceived Externalization of a Frontal Sound Source. In *Audio Engineering Society Convention 144*; Audio Engineering Society: Milan, Italy, 2018.

21. Bailey, W.; Fazenda, B. The Effect of Visual Cues and Binaural Rendering Method on Plausibility in Virtual Environments. In *Audio Engineering Society Convention 144*; Audio Engineering Society: Milan, Italy, 2018.

22. Werner, S.; Klein, F.; Mayenfels, T.; Brandenburg, K. A summary on acoustic room divergence and its effect on externalization of auditory events. In Proceedings of the International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016; pp. 1–6.

23. Hupke, R.; Nophut, M.; Li, S.; Schlieper, R.; Preihs, S.; Peissig, J. The Immersive Media Laboratory: Installation of a Novel Multichannel Audio Laboratory for Immersive Media Applications. In *Audio Engineering Society Convention 144*; Audio Engineering Society: Milan, Italy, 2018.

24. Müller, S.; Massarani, P. Transfer-function measurement with sweeps. *J. Audio Eng. Soc.* **2001**, *49*, 443–471.

25. Glasberg, B.R.; Moore, B.C. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* **1990**, *47*, 103–138. [CrossRef]

26. Kohlrausch, A.; Breebaart, J. Perceptual (ir) relevance of HRTF magnitude and phase spectra. In *Audio Engineering Society Convention 110*; Audio Engineering Society: Amsterdam, The Netherlands, 2001.

27. Holdsworth, J.; Patterson, R.; Nimmo-Smith, I.; Rice, P. Implementing a gammatone filter bank. *Annex C SVOS Final Rep. Part A* **1988**, *1*, 1–5.

28. Toole, F. In-Head Localization of Acoustic Images. *J. Acoust. Soc. Am.* **1970**, *48*, 943–949. [CrossRef] [PubMed]

29. Panzer, J.; Ferekidis, L. The use of continuous phase for interpolation, smoothing and forming mean values of complex frequency response curves. In *Audio Engineering Society Convention 116*; Audio Engineering Society: Berlin, Germany, 2004.

30. Schärer, Z.; Lindau, A. Evaluation of equalization methods for binaural signals. In *Audio Engineering Society Convention 126*; Audio Engineering Society: Munich, Germany, 2009.

31. Puckette, M. Combining event and signal processing in the MAX graphical programming environment. *Comput. Music J.* **1991**, *15*, 68–77. [CrossRef]

32. EBU. 3253: Sound Quality Assessment Material Recordings for Subjective Tests. 2008. Available online: https://tech.ebu.ch/docs/tech/tech3253.pdf (accessed on 15 September 2018).

33. Lindau, A.; Weinzierl, S. Assessing the plausibility of virtual acoustic environments. *Acta Acust. United Acust.* **2012**, *98*, 804–810. [CrossRef]

34. Satongar, D.; Pike, C.; Lam, Y.W.; Tew, A.I. The influence of headphones on the localization of external loudspeaker sources. *J. Audio Eng. Soc.* **2015**, *63*, 799–810. [CrossRef]

35. Begault, D.R.; Wenzel, E.M. Headphone localization of speech. *Hum. Factors* **1993**, *35*, 361–376. [CrossRef] [PubMed]

36.   Cubick, J.; Rodríguez, C.S.; Song, W.; MacDonald, E. Comparison of binaural microphones for externalization of sounds. In Proceedings of the 3rd International Conference on Spatial Audio, Graz, Austria, 17 September 2015.

37.   Morimoto, M. The contribution of two ears to the perception of vertical angle in sagittal planes. *J. Acoust. Soc. Am.* **2001**, *109*, 1596–1603. [CrossRef] [PubMed]

38.   Macpherson, E.A.; Sabin, A.T. Binaural weighting of monaural spectral cues for sound localization. *J. Acoust. Soc. Am.* **2007**, *121*, 3677–3688. [CrossRef] [PubMed]

39.   Baumgartner, R.; Majdak, P.; Laback, B. Modeling sound-source localization in sagittal planes for human listeners. *J. Acoust. Soc. Am.* **2014**, *136*, 791–802. [CrossRef] [PubMed]

40.   Baumgartner, R.; Majdak, P.; Colburn, H.S.; Shinn-Cunningham, B. Modeling sound externalization based on listener-specific spectral cues. *J. Acoust. Soc. Am.* **2017**, *141*, 3630. [CrossRef]